Journal of Statistical Planning and Inference 26 (1990) 131-148
North-Holland

Minimax and maximin distance designs*

M.E. Johnson

Georgia Institute of Technology, Atlanta, GA 30332, U.S.A.

L.M. Moore

Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

D. Ylvisaker**

UCLA, Los Angeles, CA 90024, U.S.A.

Received 18 October 1988; revised manuscript received 19 April 1989
Recommended by C.S. Cheng

Abstract: Beginning with an arbitrary set and a distance defined on it, we develop the notions of minimax and maximin distance sets (designs). These are intended for use in the selection-of-sites problem when the underlying surface is modeled by a prior distribution and observations are made without error. It is shown that such designs have quite general asymptotically optimum (and dual) characteristics under what are termed the G- and D-criteria. There are many examples given, dealing especially with the unit square and with k factors at two levels.

AMS Subject Classification: 62K05.

Keywords: Bayesian design; asymptotic optimality; computer experiments.

1. Introduction

The paper is intended first as a contribution to the recent work on the design of computer experiments as it has been reported in Sacks and Schiller (1987), Sacks, Schiller and Welch (1989), and Currin, Mitchell, Morris and Ylvisaker (1988), for example. At each t in a set of 'sites' T, one may run a computer program with resulting deterministic output $x(t)$ (here taken to be univariate, although one can work out some extensions at the cost of added complications). It is envisioned that the program models a complex physical system (via a set of partial differential equations, for instance) and $x(t)$ is the desired reading corresponding to the initial conditions t. Specific examples along these lines can be found in the references above. Statistics and design enter through the device of placing a prior measure on the func-

* Work performed under the auspices of the U.S. Department of Energy.
** Research supported in part by NSF Grant DMS 86-02018.

tion space $\{x(t), t \in T\}$ as representative of one's ignorance about the underlying relationship. Subsequently one decides on sites at which to run the program and analyzes the data through updating of the prior.

The 'Bayesian' philosophy suggested above is hardly new as it has surfaced in many areas long before computer output modeling, notably in geostatistics and in numerical analysis (see Diaconis (1987) for an entertaining account of its surprisingly rich history). Our results might yield some perspective to other areas as well, but we will leave many natural questions unanswered.

The basic concern here is with 'minimax' and 'maximin' distance sets. In our view these are statistical designs, i.e., places at which to observe the variable of interest. The underlying notions make sense in quite general settings, and the designs which result have some optimal or asymptotically optimal characteristics in suitable problems. The usage implicitly proposed has the effect of transferring attention to more geometric aspects of designs and of their generation.

We begin with the distance ideas in Section 2, then turn to the Bayes set-up and various criteria for distinguishing good designs in Section 3. After these preliminaries, asymptotics are introduced in Section 4 and results follow on the optimality and asymptotic optimality of particular designs. This treatment is done specifically for the case of finite T, while some attention is given to infinite T in an appendix. Our considerations are made a good deal more specific in Section 5 by focussing on some contexts, discrete and continuous, which are commonly met as experimental conditions.

2. Designs

Let T denote a set of sites and suppose there is a nonnegative function d on $T \times T$ so that

$$d(s,t) = d(t,s) \quad \text{for all } s,t \text{ in } T, \tag{2.1.i}$$

$$d(s,t) \geq 0 \quad \text{with equality if and only if } s = t. \tag{2.1.ii}$$

If one allows further that

$$d(s,t) \leq d(s,u) + d(u,t) \quad \text{for all } s,t \text{ and } u \text{ in } T, \tag{2.1.iii}$$

then (T, d) is an honest metric space. On the one hand we refer to d as a distance function, on the other hand we make scant use of (2.1.iii).

2.1. Consider subsets S of T with $\text{card}(S) = n$, n fixed. Call S^* a *minimax distance design* if

$$\min_{S} \max_{t \in T} d(t, S) = \max_{t \in T} d(t, S^*) = d^* \tag{2.2a}$$

where $d(t,S) = \min_{s \in S} d(t,s)$. If necessary one should replace min by inf, max by sup, and proceed with a notion of ε-minimaxity.

Imagine a set S^* of the above type as one solution to a franchise-store placement problem: the customer (site t) furthest removed from the nearest store location (s^*) is as close as can be managed. From a design standpoint, the thought is that $x(t)$ will be predicted well when t is near design sites and less well when t is remote from all observed sites.

Examples 2.1. (1) With ordinary distance operating on [0,1] one places n points at elements of

$$S^* = \{(2i-1)/2n, i = 1, \ldots, n\}$$

while $d^* = 1/2n$.

(2) With $T = [0,1]^2$ and Euclidean distance, for $n = 3$, a choice of S^* is

$$\{(\tfrac{1}{16}, \tfrac{1}{2}), (\tfrac{9}{16}, \tfrac{1}{2} \pm \tfrac{1}{4})\},$$

where $d^* = \frac{1}{16}\sqrt{65}$.

(3) For the same problem as in (2) but with the rectangular distance

$$d((s,\delta),(t,\tau)) = |s-t| + |\sigma - \tau|,$$

take $S^* = \{(\frac{1}{8}, \frac{1}{2}), (\frac{5}{8}, \frac{1}{2} \pm \frac{1}{4})\}$ with $d^* = \frac{5}{8}$, as one possibility.

(4) Another choice of d, suggested by developments in Section 3, is given by

$$d(s,t) = 1 - \{\min(s,t)/\max(s,t)\}^{1/2}, \quad s \text{ and } t > 0.$$

In this case, if $T = [1,2]$, the problem of placing n points yields

$$S^* = \{t_i = \exp\{(4i-2)(\ln 2/4n)\}, i = 1, \ldots, n\}$$

and $d^* = 1 - \exp\{-(\ln 2/2n)\}$.

There is commonly non-uniqueness in the choice of a minimax distance set, and one distinction which can be made between different sets will show up in a crucial way. *Remotes sites* to S^* are those t for which $d(t, S^*) = d^*$. Minimax distance sets will be assigned an *index* I^* – it is the least number of sites in S^* which are at distance d^* from some remote site. S^{**} will denote any minimax distance set with highest index. Using the franchise-store analogy, S^{**} yields a guarantee to remote customers that they have been given the greatest possible choice in alternative stores. It may of course be that S^{**} is again non-unique and indeed, every S^* in the examples above has index 1.

2.2. Once more let S be a subset of T with $\mathrm{card}(S) = n$, n fixed. Call S° a *maximin distance design* if

$$\max_S \min_{s,s' \in S} d(s,s') = \min_{s,s' \in S^\circ} d(s,s') = d^\circ. \tag{2.2b}$$

It may be that there is not enough compactness to guarantee a true maximum in which case one considers ε-maximinity. The *index* I° of a maximin set is the number of pairs of sites in S° separated by distance d°, and $S^{\circ\circ}$ will denote any maximin set with smallest index.

In (2.2b) one might say that the viewpoint of the franchise holder is emphasized as opposed to that of the customer (in 2.2.a)). One wishes to guarantee the store operator the largest 'exclusive territory' that can be managed – it extends up to a distance of at least $\frac{1}{2}d^\circ$ by the triangle inequality. Moreover the (offending) distance d° is found least often in a maximin set of smallest index. The design emphasis here suggests nonredundancy in choice of sites.

Examples 2.2. (1) Under ordinary distance on [0, 1] one again assigns an equal spacing to n points, but now these are to include 0 and 1. Here $d=1/(n-1)$ and the index is $n-1$.

(2) With rectangular distance on $[0,1]^2$, the case $n=3$ gives three boundary sites separated pairwise by $d^\circ=\frac{4}{3}$ (index 3). This compares with the distances $\frac{1}{2}$ and $\frac{3}{4}$ between the three interior points of S^* found earlier. In such cases the presence of a second criterion suggests a natural way of dealing with the non-uniqueness which is regularly found.

There is an obvious connection in 2.1 with covering problems, and a relation between 2.2 and packing problems, as these are described in Chapters 1 and 2 of Conway and Sloane (1988), or in MacWilliams and Sloane (1977). The extensive literature on these subjects contains results which are useful to us, but our outlook is often quite different from the standard one there. In setting 2.1 for instance, our interest might be in placing n spheres to cover the unit p-cube rather than in finding an efficient covering of p-dimensional space. We then encounter boundary effects, due to experimental conditions, safe in the knowledge that the space covering answers which are known can be used for asymptotic (in n) calculations and bounds. As another example, there are results known about the packing problem on 2^k which bound the number of 'spheres' of radius r which can be packed. This is an aid, but not a solution, to the problem of proper placement of n spheres as this may or may not require an efficient packing.

3. Priors and criteria of goodness

Begin with a Gaussian process X which is indexed by the set T. Take X to have mean zero, a correlation fuction ϱ, and allow that all finite-dimensional distributions of the process are nonsingular.

Suppose $\varrho(s,t)$ is a decreasing function of $d(s,t)$ where d satisfies (2.1), say $\varrho(s,t)=r(d(s,t))$. Before proceeding, it is useful to spend some time over this last assumption as we think of it. Some of our examples are straightforward: if $T=\mathbf{R}^P$ and d is Euclidean distance, one meets stationary and isotropic correlation functions; for $T=\mathbf{R}^P$ and $d(s,t)=\sum_i |s_i-t_i|$, a particular class of correlations is given through

$$\varrho(s,t) = \int \lambda^{d(s,t)}\,\mathrm{d}F(\lambda) \tag{3.1}$$

where F is a probability distribution on $(0,1)$; when $T=\{0,1\}^P$ and d is the Hamming distance, the family of correlation functions can be characterized by the inequalities

$$\sum_{j=0}^{p} r(j) \sum_{i=0}^{k} (-1)^i \binom{k}{i}\binom{p-k}{j-i_0} \geq 0, \quad k=0,1,\ldots,p \tag{3.2}$$

(there is a connection here with Krawtchouk polynomials, see Letac (1981)). A different type of example simply begins with a nonnegative $\varrho(s,t)$ and takes $d(s,t)=1-\varrho(s,t)$, with $r(z)=1-z$. A particular instance of this arises if one computes the correlation function for the Brownian motion to reach $d(s,t)=1-(\min(s,t)/\max(s,t))^{1/2}$, a distance brought up earlier to reach the d of Example 2.1(4).

For correlations depending on distance as above, we say ϱ is *local* of order d_0 provided $d_0=\inf\{d \mid r(z)=0 \text{ if } z\geq d\}$. A class of local stationary and isotropic correlation functions can be discerned from Mittal (1976), and inspection of (3.2) will reveal that there are local correlations on $\{0,1\}^P$ of order q for any $q\leq p$. In particular one can take $r(d)=\varepsilon_d$ for $d\leq q-1$, $r(d)=0$ for $d\geq q$, ε_d small enough.

Now consider the design problem of choosing a set S, $\mathrm{card}(S)=n$, at which to observe the process X. Again the references Sacks and Schiller (1987), Sacks, Schiller and Welch (1989), and Currin, Mitchell, Morris and Ylvisaker (1988) go into reasons to think of (i) X as a prior distribution for some response function on T and (ii) the problem as one of Bayesian design for eliciting this response. Throughout this section we shall suppose *T is a finite set* and thus that there are but a finite number of competing design sets S to contend with. While the minimax and maximin notions make sense for any T, more complex issues arise when T is infinite. These are deferred to an appendix lest the discussion become bogged down in taming assumptions here. As justification for considering finite T there are already 'regular' cases like $T=\{0,1\}^P$, but we also think of discretizing some region of $\mathbf{R}^P$, say, as both a technically interesting and an eminently practical way to deal with a full continuous setting.

For a fixed S, generate the conditional distribution of X given X_s, $s\in S$, and minimize one of the following by choice of S:

$$m(S) = \max_t \{\mathrm{var}(X_t \mid X_s,\ s\in S)/\mathrm{var}(X_t)\}, \tag{3.3.G}$$

$$M(S) = -\det(\mathrm{corr}\{X_s,\ s\in S\}), \tag{3.3.D}$$

$$\mu(S) = \sum_t \mathrm{var}(X_t \mid X_s,\ s\in S)/\mathrm{var}(X_t), \tag{3.3.A}$$

where corr{ } denotes the $n\times n$ correlation matrix of the designated variables. Minimizing designs are deemed G-, D- and A-optimal, respectively. (This terminology is supposed to be suggestive; it comes from design theory and more explanation will be added later.)

We will now work up to the link between (3.3.G) and (3.3.D) and the minimax and maximin distance designs of Section 2. The latter designs will be shown to have a suitable asymptotically optimal character. Our interest in local correlations is subsequently made plain when we look to establish actual optimality of the designs for particular processes (as opposed to a sequence of processes) X.

4. Asymptotics and optimality

A form of asymptotics comes about through the following device: if ϱ is a correlation function on $T\times T$, so is ϱ^k for any $k=2,\ldots$. Then allow that $k\to\infty$. Another perspective comes from thinking of T as a linear metric space. For a correlation function $\varrho=r(d)$ which decreases in d, it follows that $\varrho_\lambda=r(\lambda d)$ is another such correlation function. Take $\lambda\to\infty$. This form of scaling asymptotics can be contrasted with that used by Lim, Sacks, Studden and Welch (1988) wherein $\lambda\to 0$. (In fact they looked specifically at the case $\varrho_\lambda(s,t)=\exp\{-\lambda d^2(s,t)\}$ for Euclidean distance in $\mathbf{R}^P$.) In plain words, they deal with near-dependence while we focus on near-independence.

We do not, of course, believe that two outputs from a computer experiment resulting from very similar inputs are well-modeled by near-independence. What we do suggest is that, in the initial phase of data gathering, the sites for observation will be remote from each other since they are few in number and the space is relatively large. Designs which result from these considerations have been put to use as first-step designs in Currin, Mitchell, Morris and Ylvisaker (1988). We turn to their asymptotic properties.

4.1. *If $\varrho=r(d)$ is a correlation function and r is a decreasing function, a minimax distance design S^{**} of highest index is asymptotically* G-*optimum for ϱ^k as $k\to\infty$.*

(Asymptotic optimality is taken to mean that for any sequence of designs S_k, $(1-m_k(S_k))/(1-m_k(S^{**}))\le 1+o(1)$, where m_k is given at (3.3.G) under the correlation ϱ^k.)

Here is the sketch of a proof. The ratio of conditional variance to variance given at (3.3.G) can be written as

$$1-r^k(d(t,S))'\{r^k(d(S,S))\}^{-1}r^k(d(S,t))$$

where $r^k(d(S,t))$ denotes the n-dimensional column vector with entries $r^k(d(s,t))$, $s\in S$, for example (it should be recalled that all design matrices have been assumed nonsingular). Thus the problem might be posed as finding the maximum over S of the minimum over $t\in T$ of the expression

$$L_k(t,S)=r^k(d(t,S))'\{r^k(d(S,S))\}^{-1}r^k(d(S,t)). \tag{4.1}$$

Observe that $L_k(t,s)$ is bounded below by $\lambda_k(S)\sum_s r^{2k}(d(s,t))$ and above by $\lambda^k(S)\sum_s r^{2k}(d(s,t))$ where $\lambda_k(S)$ and $\lambda^k(S)$ are the smallest and largest eigenvalues of $\{r^k(d(S,S))\}^{-1}$. Now for any S,

$$\sum_s r^{2k}(d(s,t))$$

$$= \exp\{\ln(\operatorname{card}\{s\in S \mid d(s,t)=d(t,S)\})+2k\ln(r(d(t,S)))\}[1+\mathrm{o}(1)]. \quad (4.2)$$

It then follows from (4.2) that if S^{**} is a minimax distance set of highest index and $\{S_k\}$ is arbitrary,

$$\min_t L_k(t,S_k)/\min_t L_k(t,S^{**})$$

$$\leq \lambda^k(S_k)\min_t \exp\{2k\ln r(d(t,S_k))$$

$$+\ln(\operatorname{card}\{s\in S_k \mid d(s,t)=d(t,S_k)\})\}[1+\mathrm{o}(1)]$$

$$/\lambda_k(S^{**})\min_t \exp\{2k\ln r(d(t,S^{**}))$$

$$+\ln(\operatorname{card}\{s\in S^{**} \mid d(s,t)=d(t,S^{**})\})\}[1+\mathrm{o}(1)].$$

This last is bounded by $1+\mathrm{o}(1)$ for large k since for any S, $\{r^k(d(S,S))\}\to I$ so that $\lambda_k(S)$ and $\lambda^k(S)$ both tend to 1 as $k\to\infty$.

Remark 4.1. A weaker form of asymptotic optimality would require of S^* that

$$(1-m_k(S_k))^{1/2k}/(1-m(S^*)))^{1/2k}\leq 1+\mathrm{o}(1).$$

Then any minimax distance set would be asymptotically optimum inasmuch as

$$(1-m_k(S_k))^{1/2k}\sim\left\{\sum_s r^{2k}(d(t,s))\right\}^{1/2k}\sim\max_s r(d(t,s))=r(d(t,S_k)).$$

Remark 4.2. When the underlying process is a G-Map (see Ylvisaker (1987) for some details) then the conditional variance of X given X_s, $s\in S$ may be interpreted as the time a Markov process with killing spends at the point t before it hits the set S. If the exponential parameter θ associated with the killing time is allowed to go to infinity (so that death comes quickly), the only reduction in conditional variance stems from fast entry to the set S. Informally, a good design must resemble a minimax distance set when distance is appropriately defined. This is the set-up which initially led us to a more systematic look at these designs.

4.2. *If $\varrho=r(d)$ is a correlation function and r is a decreasing function, a maximin distance design $S^{\circ\circ}$ of lowest index is asymptotically* D-*optimum for ϱ^k as $k\to\infty$.*

This claim is verified by the following simple argument. The principal terms in the expression $1-\det(\operatorname{corr}\{X_s,\ s\in S\})$ are in the sum $\sum_{s,s'\in S} r^{2k}(d(s,s'))$. For asymptotic minimization, the largest terms must be minimized. But this means that the smallest distance appearing should be maximized and should appear as infrequently as can be managed.

Remark 4.3. The use of G- and D-optimality mimics the usage found in classical design, see Kiefer and Wolfowitz (1959), for example. There G-optimality refers to the minimization of the maximum variance of the fitted response over the design region while a D-optimum design minimizes the determinant of the covariance matrix of best estimates. The celebrated theorem of Kiefer and Wolfowitz (1960) finds these apparently different criteria leading to the same designs. Here the D-criterion is equivalent to minimizing the determinant of the conditional covariance matrix at unobserved locations, D-optimum designs are more readily obtained (advantage) and have the property (disadvantage?) that sites tend to lie toward or on boundaries.

4.3. We consider briefly A-optimality. This amounts to integrated (that is, summed) mean-squared prediction error.

As at (4.1) we see that the problem under the correlation comes down to maximizing the sum over $T-S$ of $L_k(t,S)$. Using eigenvalue bounds again, one should maximize $\sum_S \sum_{T-S} r^{2k}(d(s,t))$. Now if d_- is the smallest intersite distance, S will be asymptotically optimum provided

$$\sum_S \operatorname{card}\{t \mid t \notin S, d(s,t) = d_-\} = \max! \tag{4.3}$$

Example 4.1. For purposes of comparison, consider the integer lattice restricted to $[0,8]^2$. Settle on rectangular distance for convenience and position three sites as at (2.2) and (4.3). For (2.2.b) one might take S° to consist of $(0,3)$, $(6,8)$ and $(8,0)$ with $d^\circ = 10$ and index 1. There are 12 such designs and they include only boundary points. (An additional 16 maximin designs contain an interior point and have index 3.) A design for (2.2.a), mentioned earlier, is $S^* = \{(1,4),(5,2),(5,6)\}$. By small perturbations of rotations of S^* one finds a total of 470 designs with $d^* = 5$. Among these 470 is S° above – it is both minimax and maximin distance. Now to satisfy (4.3) one need only insure that design sites are not in the boundary and all intersite distances are at least 2 (then (4.3) has value 12). There are more than 15 000 such and these include designs like $S\{(1,1),(1,3),(1,5)\}$ and S^* (but not S°) above.

To conclude the section it is instructive to keep the setting of Example 4.1 and to give arguments for exact G- and D-optimality of minimax and maximin distance designs under specific (local) correlation functions.

Let $\varrho = r(d)$ where r is local of order 10. We need not insist that r be monotone at the moment and one can take $r(i) = \varepsilon_i$ for $i = 1, \dots, 9$, $r(i) = 0$ for $i = 10, \dots, 16$ if the ε_i are small enough, as an example. Then note that any S° is D-optimum since

$$M(S) = -|S^\circ| = -1 \le -|S^\circ| = M(S), \quad \text{for all } S.$$

Moreover the inequality is strict unless S is a maximin distance design if, for instance, one takes r to be decreasing.

Next take r to be monotone, local of order 6 and observe that if S has remote sites at a distance of 6 or more, $m(S)=1$. Thus G-optimum designs are to be found among minimax distance designs. The largest scaled conditional variance for S^* above is $1-r^2(5)$ and is obtained at $(0,0)$, a remote site of index 1 whose nearest observed site $(1,4)$ is at distance 6 from the other observed sites. It can be argued, though not with great educational value, that any of the 470 minimax distance designs has the same maximum value and is therefore G-optimum.

Such arguments can be put to further use, although real modifications become necessary when we get away from finite T. See the appendix for more details.

5. Examples

The intention here is to provide a number of illustrations of minimax and maximin distance designs so that the reader may tie these notions down. We aim for some contrasts by choice of distance and criterion. On the other hand we do not strive to present complex examples and will not provide proofs that certain designs do what they are claimed to do – indeed it may be that subtlety or two has escaped our notice.

The section begins with designs on the unit square as these can be found as well as pictured. Following such an exhibition, we present some examples of designs in the k-factor, 2-level context. The sample size n is taken from 'small' to 'large' in the settings considered. To save space we will write mM and Mm to denote minimax and maximin, respectively. We know something of the computational complexities to be found outside of these frameworks, but will not address them here.

5.1. *The unit square.* As experimental conditions the square stands in for two variables which can vary independently, each over a bounded interval. There will generally be non-uniqueness of designs in some cases, stemming only from rotation of the square, but we feel free to name particular choices in this regard.

For mM distance designs with $n=3$, and both Euclidean (e) and rectangular (r) distance, one finds

$$S_{\mathrm{e}}^* = \{(\tfrac{1}{16},\tfrac{1}{2}),(\tfrac{9}{16},\tfrac{1}{2}\pm\tfrac{1}{4})\},\quad d_{\mathrm{e}}^* = \tfrac{1}{16}\sqrt{65},$$

$$S_{\mathrm{r}}^* = \{(\tfrac{1}{8},\tfrac{1}{2}),(\tfrac{5}{8},\tfrac{1}{2}\pm\tfrac{1}{4})\},\quad d_{\mathrm{r}}^* = \tfrac{5}{8}.$$

These are depicted in Figure 1 (which suggests the logic underlying S^* by the placement of circles about design points) and Figure 2. It is to be noticed that there is little effect here due to the choice of distance and that our designs have been chosen with the same 'orientation'.

For $n=5$, the corners–center design is Mm distance with $d_{\mathrm{e}}^\circ = \tfrac{1}{2}\sqrt{2}, d_{\mathrm{r}}^\circ = 1$.

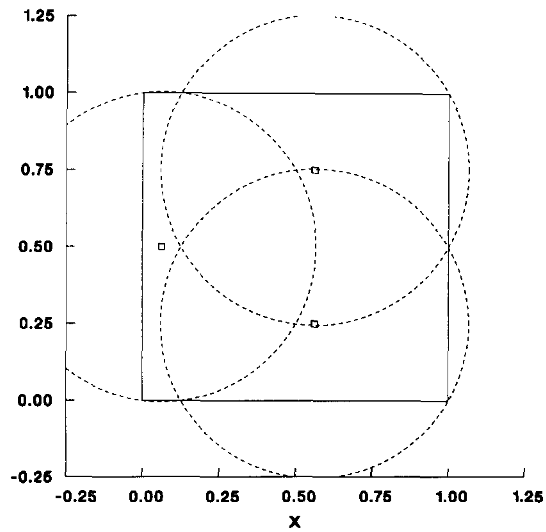

Fig. 1. Minimax Euclidean distance design for $n=3$ points in $[0,1]^2$.

Meanwhile S_e^* and S_r^* again are close to each other with

$$S_e^* = \{(0.209, \tfrac{1}{2} \pm \tfrac{1}{4}), (0.709, \tfrac{1}{2} \pm 0.353), (0.844, \tfrac{1}{2})\},$$

$$S_r^* = \{(\tfrac{3}{16}, \tfrac{1}{2} \pm \tfrac{1}{4}), (\tfrac{11}{16}, \tfrac{1}{2} \pm \tfrac{3}{8}), (\tfrac{13}{16}, \tfrac{1}{2})\}$$

and $d_e^* = 0.326, d_r^* = \frac{7}{16}$. See Figures 3 and 4.

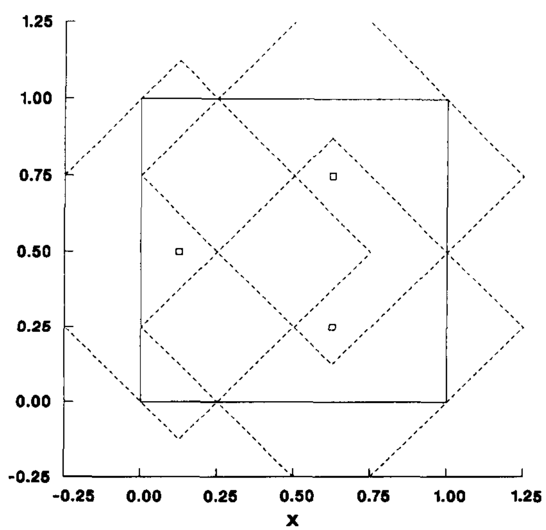

Fig. 2. Minimax rectangular distance design for $n=3$ points in $[0,1]^2$.

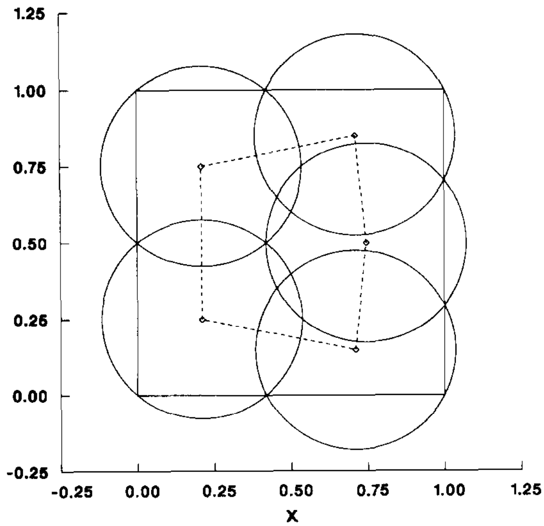

Fig. 3. Minimax Euclidean distance design for $n=5$ points in $[0,1]^2$.

When $n=7$, mM design points show on the boundary for the first time and we find

$$S_{\mathrm{e}}^* = \{(\tfrac{1}{2},\tfrac{1}{2}),(\tfrac{1}{2},\tfrac{1}{2}\pm\tfrac{1}{2}),(\tfrac{1}{3}-\tfrac{1}{12}\sqrt{7},\tfrac{1}{2}\pm\tfrac{1}{4}),(\tfrac{2}{3}+\tfrac{1}{12}\sqrt{7},\tfrac{1}{2}\pm\tfrac{1}{4})\},$$

$$\mathrm{d}_{\mathrm{e}}^* = \tfrac{1}{6}(\sqrt{7}-1) = 0.274,$$

$$S_{\mathrm{r}}^* = \{(\tfrac{1}{2},\tfrac{1}{2}),(\tfrac{5}{6},\tfrac{5}{6}),(\tfrac{1}{6},\tfrac{1}{6}),(0,\tfrac{2}{3}),(\tfrac{2}{3},0),(1,\tfrac{1}{3}),(\tfrac{1}{3},1)\}, \quad d_{\mathrm{r}}^* = \tfrac{1}{3}.$$

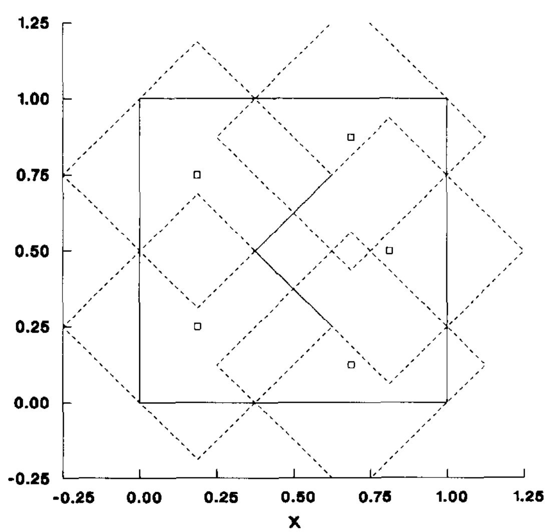

Fig. 4. Minimax rectangular distance design for $n=5$ points in $[0,1]^2$.

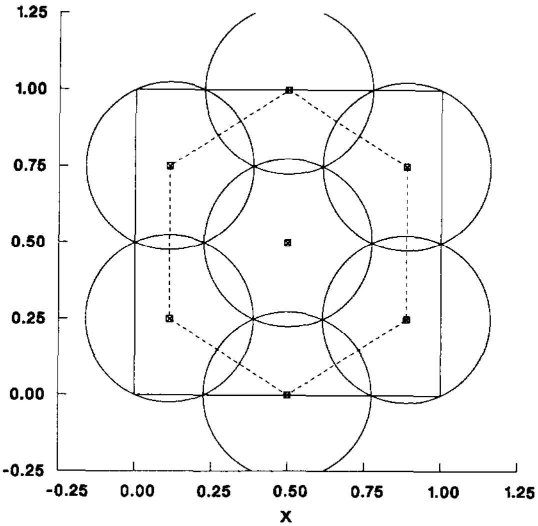

Fig. 5. Minimax Euclidean rectangular distance design for $n=7$ points in $[0,1]^2$.

These sets are visible in Figures 5 and 6. In this case Mm designs have been exhibited in Figures 7 and 8 (a and b). Mm design points and distances for $n=7$ are as follows.

$$S_e^{\circ}=\{(0,0.0941),(0,0.09059),(1,0.2343),(1,0.7657),(0.3430,\tfrac{1}{2}),$$
$$(0.5230,0),(0.5230,1)\},\quad d_e^{\circ}=0.5314;$$
$$S_r^{\circ}=\{(0,0),(0,1),(1,0),(1,1),(\tfrac{1}{6},\tfrac{1}{2}),(\tfrac{5}{6},\tfrac{1}{2}),(\tfrac{1}{2},\tfrac{5}{6})\},\quad d_r^{\circ}=\tfrac{2}{3}.$$

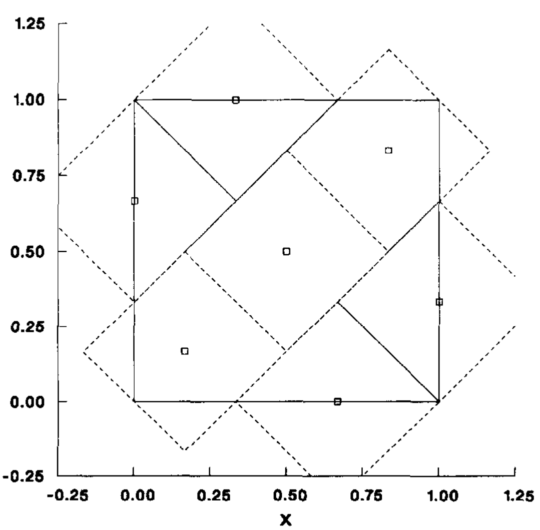

Fig. 6. Minimax rectangular distance design for $n=7$ points in $[0,1]^2$.

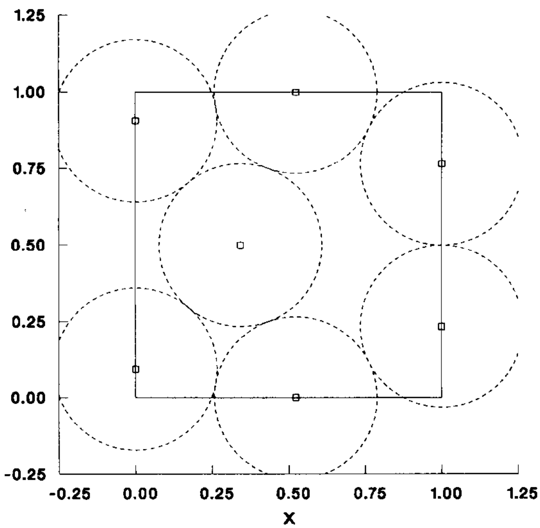

Fig. 7. Maximin Euclidean distance design for $n=7$ points in $[0,1]^2$.

Note in Figure 8 (a and b) that there are eight points shown and a seven point Mn design follows in each case be deleting any point.

Finally we present in Figures 9 and 10 some designs we have found for $n \cong 127$.

5.2. 2^k. We will give a number of designs of mM and Mm distance type for the situation of k factors each at two levels. Hamming distance will be in use and the notation will identify by letter those factors to be run at high level, where (1) denotes the all low-level design. We will consider in particular efficient covering designs.

If $k=n=2$, the design $\{(1), \text{ab}\}$ is both mM and Mm with $d^*=1$ (and $d^\circ=2$). If n is then taken to be 3, all sets are mM and Mm with $d^*=d^\circ=1$. This is a general feature for large n relative to the efficient covering number of k – there are many designs with the desired property.

Next take $k=3$ and $n=2$. The design $\{(1), \text{abc}\}$ is both mM and Mm with $d^*=1$ and $d^\circ=3$. With $n=3$ one finds 24 mM designs consisting of a foldover pair and any other point. There are eight Mm designs with $d^\circ=2$, and none of these is mM. Finally $n=4$ gives but two Mm designs $(d^\circ=2)$:$\{(1), \text{ab}, \text{ac}, \text{bc}\}$ and its foldover. The bulk of all designs are mM.

Let $k=4$. There are 40 four-point designs which are mM with $d^*=1$. The two types are exemplified by $\{(1), \text{ac}, \text{bd}, \text{abcd}\}$, two foldover pairs, and $\{(1), \text{ab}, \text{acd}, \text{bcd}\}$. The first of these types has $d^\circ=2$ and is Mm, the second has $d^\circ=1$ and is not. The Mm distance of $d^\circ=2$ can be preserved up through $n=8$, as in $\{(1), \text{ab}, \text{ac}, \text{ad}, \text{bc}, \text{bd}, \text{cd}, \text{abcd}\}$.

For $k=5$, the efficient covering number for $d^*=1$ is $n=7$. There are 320 mM designs. To achieve $d^*=2$ one requires only a single foldover pair, so $d^*=2$ for

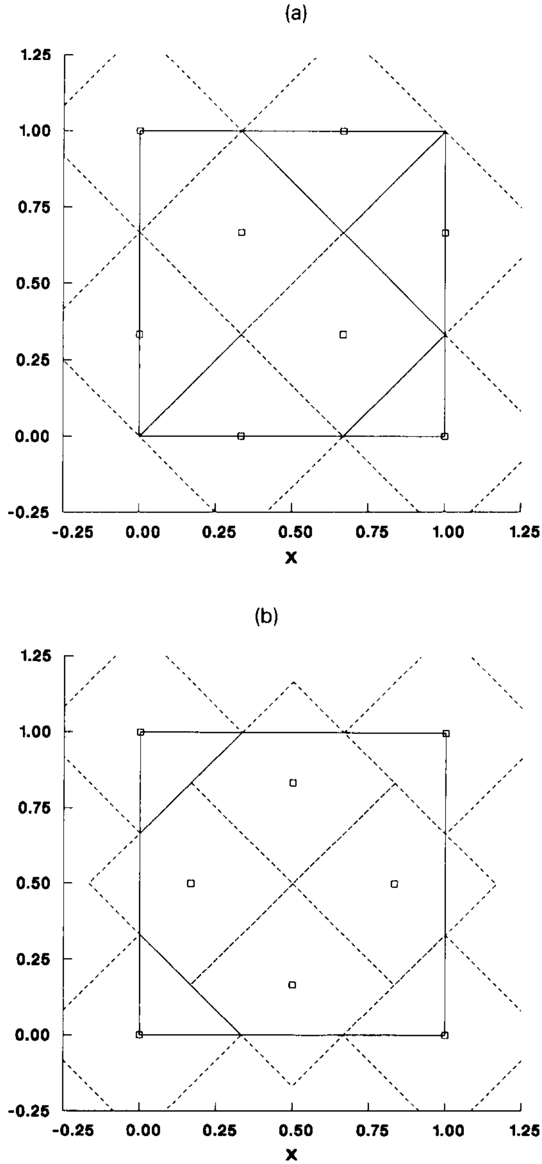

Fig. 8 (a and b). Maximin rectangular distance designs for $n=8$ points in $[0,1]^2$. Note that omission of any point leaves maximin designs for $n=7$ points.

$n=2$, 3, 4, 5, and 6. It is possible to find a Mm design with $d^{\circ}=2$ for n up to 22, with $d^{\circ}=3$ for n up to 4.

Finally let $k=6$ and note that the efficient covering number for $d^*=1$ is $n=12$. Here there are at least two nonisomorphic types: a foldover design such as

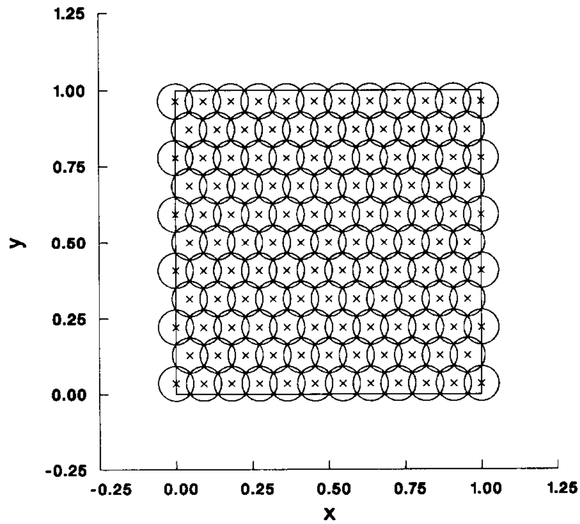

Fig. 9. Minimax Euclidean distance design for $n = 127$ points in $[0,1]^2$.

{(1), ab, ac, bcd, bce, ade, abcdef, cdef, bdef, aef, adf, bcf}; a nonfoldover design such as {(1), a, bcd, e, abce, abde, bcf, bdf, acdf, abef, cdef, acdef}.

Appendix

We consider the extension of the exact optimality arguments used at the end of

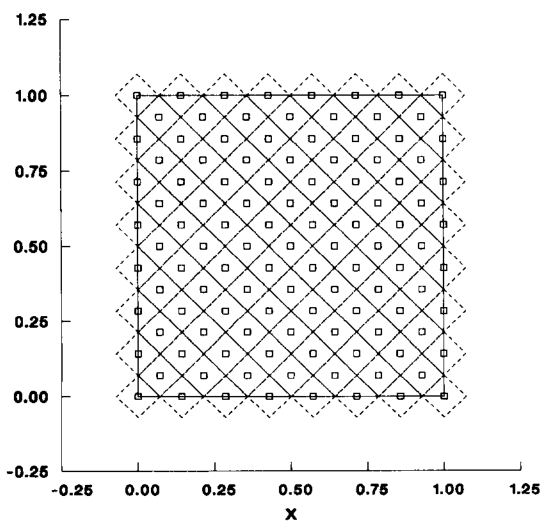

Fig. 10. Minimax rectangular distance design for $n = 127$ points in $[0,1]^2$.

Section 4 to the case that T is a general finite set. Secondly we investigate particular infinite T, chosen for convenience, concerning both exact and asymptotic D-optimality. Some remarks are added about the more difficult nature of G-optimality in such contexts. Throughout this discussion, $\varrho = r(d)$ and r is decreasing with distance. When necessary, r is assumed to be continuous.

A.1. T is finite

A.1.1. Let S° be a maximin distance set with minimum intersite distance d°. Then if ϱ is local of order d°, one finds exactly as before that $M(S^\circ) = -1 < M(S)$ for any non-maximin distance design S.

A maximin distance set S is D-*optimum for appropriate* ϱ.

A.1.2. (If S is not minimax distance (d^*) and ϱ is local of order $d+$ (where $d+$ denotes the next smallest intersite distance past d^*) then $m(S) = 1 > m(S^*)$ for any minimax distance S^*. Index does not play a direct role.

For suitable ϱ *the* G-*optimum design is to be found among minimax distance sets.*

A.2. T is infinite

It will be convenient to think of a case such as $T = [0, 1]^P$ equipped with Euclidean distance. Portions of what is then true might be carried over to other cases according to their topological structure.

A.2.1. First note that the maximin distance d° exists and that there are maximin distance sets S° (consequently one of lowest index). To see this suppose $q^P < n \leq (q+1)^P$ and take S to consist of all points of the form $(i_1/q, i_2/q, \dots, i_p/q)$, $i_j = 0, 1, \dots, q$. Here the minimum intersite distance is $1/q$, hence $d^\circ \geq 1/q$. Now the collection of all n-point sets in $[0, 1]^P$ with minimum interpoint distance $\geq 1/q$ is compact. Therefore $\min_{s, s' \in S} d(s, s')$ achieves its maximum since it is a continuous function of S.

It is easy to see that *a set S is* D-*optimum for a local correlation of order d if and only if S is maximin distance.*

It is also true that *a maximin distance* S° *of lowest index* q° *is asymptotically* D-*optimum for* ϱ^k *as* $k \to \infty$.

However, one should argue away the possibility that interpoint distances growing arbitrarily close to 0 (correlations growing close to 1) cause difficulty. Accordingly, let

$$\Psi = \left\{ S \mid r\left(\min_{s, s' \in S} d(s, s') \right) = r(d(s_-, s'_-)) > (1+\varepsilon)\, r(d^\circ) \right\}.$$

Then under ϱ^k, and using suggestive notation,

$$\inf_{\Psi}(1-M_k(S)) = \inf_{\Psi}[1-M_k(\{s_-,s'_-\})M(S-(s_-,s'_-)\mid s_-,s'_-)]$$

$$\geq \inf_{\Psi}(1-M_k(\{s_-,s'_-\})M(S-\{s_-,s'_-\}))$$

$$\geq r^{2k}(d(s_-,s'_-)) > (1+\varepsilon)^{2k}r^{2k}(d^\circ) > qr^{2k}(d^\circ).$$

Thus $\inf_{\Psi}(1-M_k(S))/qr^{2k}(d^\circ)(1+\mathrm{o}(1)) > 1+\mathrm{o}(1)$. Fix attention on those S with $r(\min_{s,s'\in S} d(s,s')) \leq (1+\varepsilon)r(d^\circ)$ or, $\min d(s,s') > d^\circ$. The argument can then proceed to a conclusion as it did in Section 2.

A.2.2. Minimax distance sets S^* exist. To see this, let S be chosen from $I=[0,1]^P\times\cdots\times[0,1]^P$ according to some fixed ordering, but allow points to be replicated. I is compact and $\max_t d(t,S)$ is a continuous function of S, hence achieves its minimum. Thus (2.2.a) admits an answer.

As one switches attention to the criterion (3.3.G) however, it may be that information grows, even as interpoint distances shrink to zero. (This is related to the difficulties which arise in simple design problems when, for example $T=[0,1]$ and X has derivatives, see Sacks and Ylvisaker (1970) for some discussion.) Here one can avoid these issues, which ultimately involve the structure of X in a complex way, by postulating a resolution distance d_-: consider only those designs with a minimum interpoint distance of at least d_-. In effect, one considers sets S which are minimax distance only over this smaller (and compact) collection (the argument given above still serves well enough to guarantee their existence). The main benefit of reducing the set of permissible designs though is this: G-optimum designs exist for continuous r, since $\mathrm{Var}\{X_t\mid X_s,\ s\in S\}/\mathrm{Var}\{X_t\}$ is jointly continuous in t and S.

The proposal now is to take minimax distance designs as approximately G-optimum ones. Our arguments in favor of this are weaker than those given in support of maximin distance designs, as can already be seen in Section 4 with regard to uniformity. It is not worthwhile to summarize the whole situation but we note here that *given any asymptotically* G-*optimum sequence* $\{S_k\}$ *for* ϱ^k, *there is a subsequence which convergences to a minimax distance design.*

The eigenvalue argument used at (4.2) works here as well since $\lambda_k(S)$ and $\lambda^k(S)$ converge uniformly to 1 over the (compact) set of designs in question. This reduces attention to $Q_k(t,S)$ where

$$r^{2k}(d(t,S)) \leq Q_k(t,S) = \sum_s 2^{2k}(d(t,\,)) \leq nr^{2k}(d(t,s)).$$

In turn one notes the inequalities

$$r^{2k}\left(\max_t d(t,S)\right) \leq \min_t Q_k(t,S) \leq nr^{2k}\left(\max_t d(t,S)\right).$$

Now if S^* is a minimax distance design and $\{S_k\}$ is asymptotically G-optimum for ϱ^k,

$$nr^{2k}\left(\max_t d(t,S_k)\right)\Big/ r^{2k}\left(\max_t d(t,S^*)\right)$$

$$\geq \min_t Q(t,S_k)/\min_t Q(t,S^*) \geq 1+\mathrm{o}(1).$$

The sequence $\{S_k\}$ has convergent subsequences and the last inequality shows any limit, S_0 say, must be minimax distance.

Acknowledgements

The authors would like to thank Ching-Shui Cheng, Toby Mitchell, Peter John and A. Juan for useful comments. This work has benefited especially from the workshops on design sponsored by the NSF.

References

Conway, J. and N. Sloane (1988). *Sphere Packings, Lattices and Groups.* Grundlehren der Wissenschaften Vol. 290, Springer-Verlag, New York.

Currin, C., T. Mitchell, M. Morris and D. Ylvisaker (1988). Bayesian design and analysis of computational experiments. ORNL Report 6498.

Diaconis, P. (1987). Bayesian numerical analysis. In: S. Gupta and J. Berger, Eds., *Fourth Symposium on Statistical Decision Theory and Related Topics.* Springer-Verlag, Berlin–New York.

Kiefer, J. and J. Wolfowitz (1959). Optimum design in regression problems. *Ann. Math. Statist.* **30**, 271–294.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12**, 363–366.

Letac, G. (1981). *Problemes Classiques des Probabilite sur Couple de Gelfand.* Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Lim, Y., J. Sacks, W. Studden and W. Welch (1988). Optimum designs for highly correlated computer experiments. Manuscript.

MacWilliams, F. and N. Sloane (1977). *The Theory of Error-Correcting Codes* 1 & 11. North-Holland, Amsterdam.

Mittal, Y. (1976). A class of isotropic covariance functions. *Pacific J. Math.* **53**, 517–538.

Sacks, J. and S. Schiller (1987). Spatial designs. In: S. Gupta and J. Berger, Eds., *Fourth Symposium on Statistical Decision Theory and Related Topics.* Springer-Verlag, Berlin–New York, 385–399.

Sacks, J., S. Schiller and W. Welch (1989). Designs for computational experiments. *Technometrics* **31**, 41–47.

Sacks, J. and D. Ylvisaker (1970). Statistical design and integral approximation. *Proc. Twelfth Biennial Seminar of the Canad. Mathematical Congress,* Montreal, 115–136.

Ylvisaker, D. (1987). Prediction and design. *Ann. Statist.* **15**, 1–19.